INTERNATIONAL DESIGN CONFERENCE - DESIGN 2016
Dubrovnik - Croatia, May 16 - 19, 2016.

DETERMINING WORK FOCUS, COMMON LANGUAGE, AND ISSUES IN ENGINEERING PROJECTS THROUGH TOPIC PERSISTANCE

Snider, C., Skec, S., Gopsill, J. A., Hicks, B. J.

Keywords: knowledge management, project monitoring, project health

1. Introduction

As with the modern world, engineering projects have evolved significantly over the past few decades with many benefits to productivity and capability. However, in tandem with this change has come a trend towards larger scale and complexity, with a single project now potentially involving many thousands of engineers, working on tens of thousands of systems and components, and generating millions of documents and communications (Watson 2012). This complexity (Earl et al. 2005) and risk (Chapman & Ward 1996) inherent even in smaller projects has great potential to cause difficulty, including delay, cost over-run and reduced quality (Xia & Lee 2004), with growing scale only exacerbating issues in project understanding and control (Floricel & Miller 2001).

One potential approach to addressing this issue is through increased capability in project monitoring (Snider et al. 2015), which has potential to stimulate successful project delivery (Wynn and Clarkson 2009) by, for example, providing feedback about process efficiency and effectiveness (O'Donnell & Duffy 2002). Typically, project monitoring has to date relied on determining the "iron triangle" of project management - time, cost, and quality (Toor & Ogunlana 2010). While this generates highly valuable information, it also often neglects the importance of the dynamic nature of enigneering projects and of varying context (Snider et al. 2015; Engwall 2002); what is important in one project is not necessarily as important in another, and importance can vary over time. Instead of static indicators, there is argument that monitoring should be based on indicators that provide reliable and regular estimation of the current project situation and its features, and can highlight potential difficulties in advance. Using real-time monitoring of a broad spectrum of project features, there is potential to quickly identify and correct issues (Durand 2014), influence project behaviour, and respond to certain project situations (Bendoly 2014). For this to be viable, however, there is a need to gather relevant, high quality, and consistent data - requirements that greatly increase the challenge of employing real-time monitoring in an organisational context.

Within engineering design research, the approach to data gathering has often relied on such means as activity logbooks, retrospective interviews and questionnaires, ethnographic observation, and protocol study. These each provide highly valuable results but, due to difficulty in large-scale implementation, struggle to provide real-time information at the required level of activity granularity. Following a more recent trend of data gathering through digital technologies (Thoring 2015), this paper looks to the digital assets (such as spreadsheets, CAD models, FEA models) produced by engineers within their every day work as a data source. Such an approach has recently shown capability in determining activity patterns (Gopsill et al. 2015), and communication patterns (Jones et al. 2015), with potential for significant further analysis through analysis of asset meta-data and content.

Through this form of data gathering, it is possible to take a new perspective on approaches to analysis of engineering activity as occurs in design research. Instead of the use of shorter-term experiments (Cash et al. 2013; Stempfle & Badke-Schaub 2002) or longer term observations that are researcher-led (see Hales 1987) or participant-led (see Robinson 2010), monitoring through digital asset analysis has potential to provide continuous long-term data that reliably represents the project from which it was developed, without requiring any intervention or input from a researcher during the gathering process.
This paper presents one such approach to project monitoring, through the analysis of emails. Emails form one of the key communication methods within engineering (Wasiak et al. 2010), particularly within distributed teams, and are increasingly a key element of personal task management (Whittaker et al. 2006) and route to understanding social interaction, content formation, and user effectiveness (O'Kane 2007). Accordingly, by analysis of email there is potential to generate information about many aspects of engineering projects.
This paper presents one approach in this thinking - analysis of the topics discussed within email to generate understanding about the project and its progress. By classifying and studying the topics discussed throughout a 142-week project based within a single company in the marine engineering sector, this paper presents an approach to automatically determining the focus of work in real-time, through detection and analysis of the topics that are being discussed in email. This has the potential to aid project managers by increasing their understanding of activity in their projects and its subject, monitoring progress, and identifying issues quickly; and aiding researchers through a new approach to large-scale, continuous, automatic analysis of engineering projects.

2. Identifying Topics

In this work, there is no prior knowledge of the project available that can be used to generate a list of suitable topics to monitor. Further, it has been shown that workers develop their own shared parlance during a project (Hill et al. 2001). Hence, the pre-definition of topics to and monitor is currently unfeasible - it is instead necessary to directly detect topics of conversation from the text itself.
In textual data, all topics must stem from individual or groups of terms within the email corpus. As the emails contain the natural written language of the workers, there is potential for any term or group of terms to form the name of, or refer to individual topics. The challenge is then to identify which terms refer to potentially meaningful topics, and which purely form working language.
There are a number of approaches to topic identification including pure term frequency (Luhn 1957), capitilsation of terms (Gruhl et al. 2004), commonly occurring sequences of terms (Gruhl et al. 2004), and words that frequently appear within a single paragraph or certain spacing (co-word analysis) (Jones et al. 2015). While these are all appropriate in certain cases, there are issues with their application herein. There are many words (such as "and", "the", "is") that are frequently used but do not represent distinct topics. As the frequency of all terms within an engineering project is not known, the data cannot be filtered for high frequency but irrelevant terms except for common stopwords, and as such techniques looking at pure frequency of terms or frequency of sequences of terms cannot be used. It can also not be assumed that topic terms will be capitilised. Finally, while co-word analysis has shown capability within similar datasets, it produces a collection of terms for each topic that change significantly both in content and semantic meaning over a project lifespan, and consequently it is not straightforward to use this method for tracking of a single topic over time. Accordingly, this paper has used an alternative method - term frequency cumulative inverse document frequency (TFCIDF) - to identify potential topic terms.

2.1 TFCIDF method

The TFCIDF approach aims to highlight individual words or sequences of words within the emails that are important to the corpus as a whole, and is used for topic identification both within research (Gruhl et al. 2004) and industry. It identifies these important words by comparing the number of emails containing each term in a given timespan (single week, in this case), to the average number of emails containing the term in each previous timespan. This is represented by the equation:

$$tfcidf(i) = (i-1)tf(i) \,/\, \sum_{j=0}^{i-1} tf(j) \quad (1)$$

Where (tfcidf) is the importance of the term, (i) is the current timespan, and (tf) is the term frequency. Thresholds can then be set for required values of (tfcidf) and (tf) above which terms are recorded as potential topics. This method sets two assumptions on terms as potential topics:

- The term has appeared a certain number of times within the given timespan, and therefore is a term that is used frequently within that timespan - determined by (tf). Hence assuming that topic names and subjects will be more frequently used than every day language.
- The term has been used a certain multiple higher than its usual use, and therefore represents usage above that which is commonly expected for that term in the given context - determined by (tfcidf). Hence assuming that topic names and subjects are not terms that are commonly used in all language over the whole project.

Following extraction, the list of terms that pass both criteria are parsed by a human to select those that are most meaningful as topics, and those that are likely to be common language in the specific context. TFCIDF was applied to the dataset using a threshold for tfcidf(i) of 3 and tf(i) of 10, as has been used in other research (Gruhl et al. 2004). The algorithm was applied for all single terms, all bi-grams (two word pairs), and all tri-grams (three words) within the data. Following, the lists of extracted terms were parsed to find those that represented the most meaningful topics, in line with the process found within literature (Gruhl et al. 2004). Summary terms and example topics are presented in Table 1.
These topics then form the basis for analysis - by tracking patterns in their occurrence through the project, information about work activities and events within the project can be implied.

Table 1: Terms selected as topics

	Total Unique	Number Extracted	Number Selected	Examples
Single terms	26,010	340	62	Cargo, transformer, drawing, vessel, pump, outfitting, requirements, specification, software, ProjectAA, CompanyAA
Bi-grams	66,097	232	53	Design department, serial link, project implementation, propulsion motor, cold ironing, main switchboard, Project AA, Company AA
Tri-grams	27,510	80	34	Outfitting design department, lotus notes release, fuel gas line, key exchange boxes, torsional vibration analysis, LV short circuit

3. Topic Occurrence and Persistence

The aim of this work is to investigate the occurrence of topics, and their persistence through the project timeline, in order to generate understanding of the activities of workers and the occurrence of specific events. In all analysis, terms are analysed individually rather than categorised. This allows a detailed understanding to be formed of specific topics, and is possible due to the high quantity of email data. While there is likely value in categorisation and detection of topic groups - for example, detection of managerial-type topics and technical-type topics from lower-level specific topics - this removes a layer of detail from analysis. For this reason it is here noted as further work.
Here topic occurrence is calculated against its typical use - occurrence is high when it is being discussed more than may be expected for that topic as an average. This recognises that different topics can have varying importance within different project contexts - what is common in one may not be common in another. In other words, it may be unremarkable that certain topics are being discussed frequently, due to the nature of the project, but remarkable that other topics are being discussed at all. By identifying against typical usage in the project context, each term can be highlighted when its occurrence is higher than may be expected. Topic persistence is then the ongoing relatively higher occurrence of a topic over the project lifespan, either consecutively or intermittently.

Identifying Persistence and Usage Episodes

Higher occurrence is detected through comparison of current usage of a term to a prior normal. During periods in which the current average is higher, the term is said to have high occurrence. The quantity, location in process, and duration of these episodes can be studied to reveal information about worker focus. This requires calculation of current usage, and of prior usage.

As the current-usage calculation must represent current term use, it is calculated as a rolling average of occurrence during the previous two weeks, proportional to the total number of emails sent within the same timespan. This 2-week timespan has been chosen as it provides a sufficiently short window for this project and for analysis, but should be tuned depending on the requirements of the manager performing analysis, the data available, and the project length.

When detecting high occurrence, it is necessary to specify a period for the rolling average of prior use of each term against which current usage can be compared. This period must be specific to the individual project (valid in the specific project context), and specific to the point in the project at which measurement occurs (valid through the changing foci of the project as it develops).

As illustrated by Fig 1, the duration of this longer-term rolling window over which the prior average is calculated has a significant impact on the detection of high occurrence and topic persistence episodes. In all sub-figures, the recent usage is shown by the solid line. Part A shows topic persistence episodes (as denoted by shaded areas) for a whole-project global average, part B for a cumulative average up to that week, and part C for a rolling 2-month window prior to each week.

When using a whole-project global average, the evolution of term usage characteristics in the project is ignored. For a term that grows in use, the increased usage in later process stages raises the global average sufficiently to ignore usage in the early process stages. As a result, the method does not detect high occurrence in early stages even though discussion may be signficant or a-typical. In addition, this method can only be applied once the global average for the whole project can be found (ie. once the project is complete), and so is not useful for real-time monitoring.

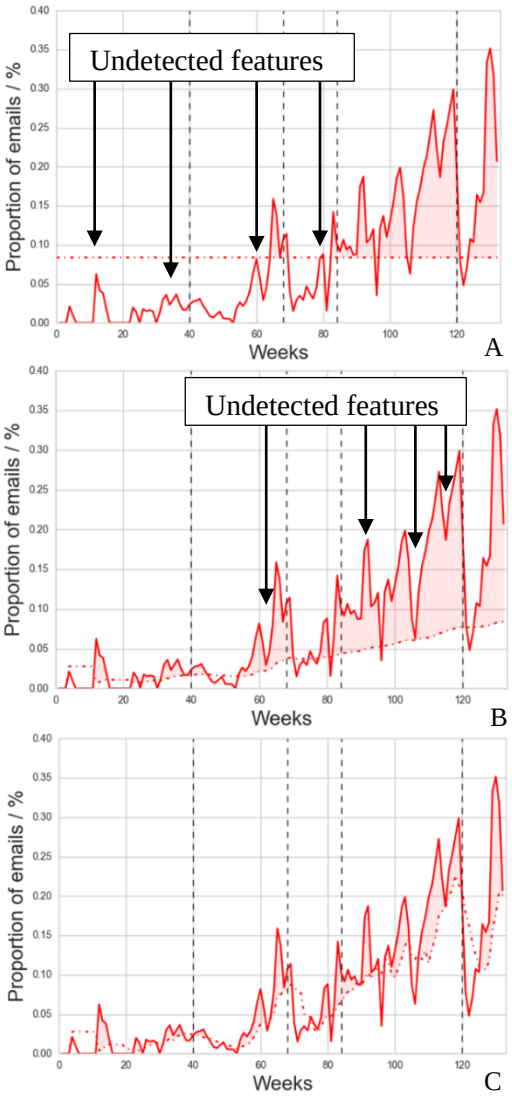

Fig 1: Effect of long-term usage calculations for the term "software"

When using a cumulative average, where the prior average is calculated from all usage up to that point in time, the value changes in tandem with term usage through time in the project. However, the high volume of data and large numbers of emails reduce the relative difference from week-to-week, particularly in later-stages when a high number of emails are sent. As a result, there is a signficant lag and important changes in usage level may be missed or established too late to be useful.

When using a 2-month rolling window for the prior usage patterns, the method accounts for changes in general usage through the project process, and is sufficiently responsive to detect short-duration changes in term usage week-by-week. As a result, it is able to quickly provide information to managers about changes in focus of their workers. Additionally, this method can be applied in real-time, and tuned to a longer or shorter long-term time window as is suitable for each individual project.

High topic occurrence is defined mathematically by the following, where a topic is said to have high occurrence at a given time (i) if occurrence (O(T)) >= 1; (tc) and (ec) are current term and email frequency, (tp) and (ep) are prior term and email frequency, (l) is timespan - 2 weeks for current and 8 weeks for prior, (f) is frequency, and (T) is each term.

$$O(T_i) = \frac{\mathrm{Ft}_c(T_i)}{\mathrm{Fe}_c(T_i)} \cdot \frac{\mathrm{Fe}(T_i)}{\mathrm{Ft}_p(T_i)} \quad ; \quad F = \sum_{j=i-l}^{i} fT_i \tag{2}$$

Once high occurrence for each topic has been found, episodes in the process in which topics were persistent can be identifie - any period of time in which (O) > 1 for one or more consecutive timespans. The points at which a term is persistent indicate high usage relative to typical usage at that point in time, and are proposed to indicate to which areas workers are paying particular attention.
In addition, patterns in persistence are thought to inform about each topic and its occurrence in the project. For example, highly persistent topics form the core themes and general chatter of the project, while short lived topics may indicate either short periods of focused work, or occurrence of potential issues that suddenly appear and must be given high attention to solve quickly.
It should be noted that the values for current and prior usage timespan have been selected here due to observed suitability to the data. In practice, values should be optimised for each individual project and the analysis that the manager wishes to perform - ie. should a 1-day recent window and 10-day longer term provide useful output, they should be used instead of those used here. This paper continues by presenting the results of analysis of topic persistence within the data set, and interpretations thereof.

4. Results

By determining topic occurrence and mapping with respect to the timeline of the project, each topic as discussed by the workers can be studied in detail. In this work these topics are presented and discussed in two ways. First in the generation of a detailed understanding of the typical work focus throughout the engineering process; second through the analysis of when in the project timeline particular topics are persistent, or are perhaps being neglected. Data for each term is mapped against engineering process stages throughout, as defined by workers within the project as data was collected. Summary statistics are presented in Table 2. Note that as stage boundaries are often fluid in nature, a one week overlap was assumed in each stage of the engineering process in all analysis.

Table 2: Email Summary Data

Process Stage	Number of emails	Period (Weeks)	Emails per Week
Whole Process	10249	1 - 142	77.6
Specification	1729	1 - 40	43.2
Manufacture	3338	40 - 68	119
Sub-system Testing	2073	68 - 84	130
Assembly	3061	84 - 120	85.0
Testing	774	120 - 132	39.2

4.1 Topic Occurrence

The occurrence of each term can be determined by equation (2) and mapped as seen in Figure 2. When mapped for each term, this gives a detailed picture of the usage of the terms through process stages. Solid lines indicate current use, dash-dot lines indicate prior, shaded areas denote O(T) >= 1.
The topics presented in Fig. 2 demonstrate some common patterns found amongst topics. Those that are very common, such as company or project names (see part A) display a high long term average and are persistent throughout the project. This can be seen by the topic "CompanyBB", which appears suddenly and maintains high levels of usage from week 85. At this point, the company performing the project changed their name, with "CompanyBB" representing the new name. There are also topics that are frequently persistent and consistent but with far lower usage, such as "control" (part A). These potentially form common, core topics for the project and its subject matter. Other topics, such as "project planning" (part B) are more periodic in nature (see repeated peaks between week 0 and 40), with appearances likely as a result of work in individual stages. The regular periodicity of "project planning" perhaps represents work prior or post-planning meeting during the specification stage. Other terms are more sporadic in appearance and persistence, such as "valve" (part B) which is prominent in manufacture and with generally rising prominence through later-stages, and "signal light columns" which appears for only a very brief period in final testing. These sporadic appearances could represent times in process when these topics are more relevant or, for short-lived but prominent topics, the discussion of an issue that must be dealt with quickly.

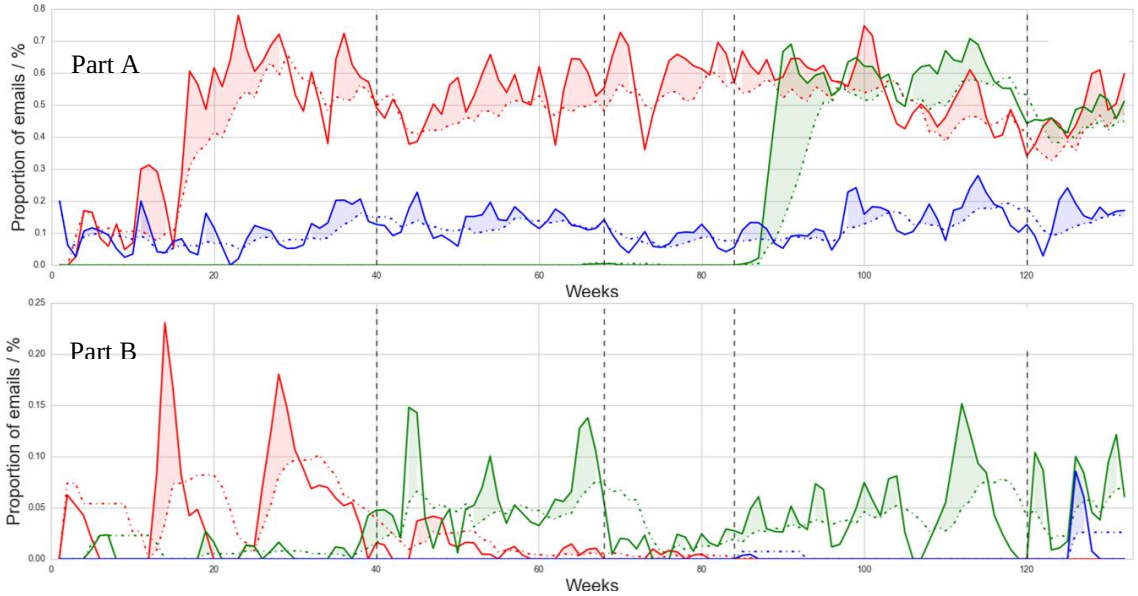

Fig 2: Occurrence for 6 terms: Above - ProjectAA (red), CompanyBB (green), "control" (blue); Below - "project planning" (red), "valve" (green), "signal light columns" (blue)

Table 3: High and low persistence topics

Persistence	Num.	Topics
> 60%	11	CompanyAA, ProjectAA, division, offshore division, drawing, software, cargo, systems
40 - 60%	55	Engine, compressor, valve, spares, protection, propulsion motor, transformer, hardware, breaker
20 - 40%	35	Flow control, change proposal, cargo handling, CompanyBB, EPS meeting, flow rate, purging
10 - 20%	21	Project planning, cargo piping, pressure transmitter, fuel gas line, spare part list, shipbuilding division
0 - 10 %	27	Lv short circuit, bus tie breakers, 20v secondary transformer, spray pipe, gcu interface, gas purging

	Mean Persistence
Upper Quartile	58.4 %
Inter-Quartile Range	33.9 %
Lower Quartile	7.25 %

Nominally, higher occurrence topics tend to be company or project names, or very general topics that seem core to the project under way. For example, "engine", "compressor", "valve", CompanyAA" in Table 3 are all high occurrence and persistence terms within the upper quartile. For a manager, the appearance of these topics may provide a general description of the core themes of the project, and may also confirm that work if occuring on these core themes. This allows a detailed understanding of work in a given context to be generated, which can be used for monitoring purposes, classification of project type, or for more detailed analysis of activity focus and sequence.

Conversely, lower occurrence topics tend to be far more specific - see "lv short circuit", "spray pipe", "gas purging", "20v secondary transformer" in Table 3. These represent topics that have a very short lifespan in the project but, due to their identification by the TFCIDF measure, should still be considered important. They may represent highly focused areas of work, or may represent topics of issues in the project, when a high amount of work was needed in a short time to diagnose and rectify. Identification of these topics may then aid a manager in understanding the areas of the project that require particular focus or greater resource and thus allocate resources more effectively, or categorise areas in which issues are more frequent.

In all cases, as discussed in Section 5, the interpretation of topics is a matter for a manager working within the correct context. The appearance and patterns of each topic may be interpreted in a number of ways, and it is for a knowledgeable observer to judge whether appearances observed are positive or negative, and whether intervention is required as a result.

4.2 Topic Persistence Within and Across Process Stages

By studying more closely the periods over which topics are persistent (here termed episodes), particularly in comparison to known process stages and boundaries, more detail of the actual work on specific topics can be formed, in context of their place in the entire project process. An episode occurs when current activity on a topic is higher than prior activity, and represents more focused work in a specific area than has typically been occuring up to that point.

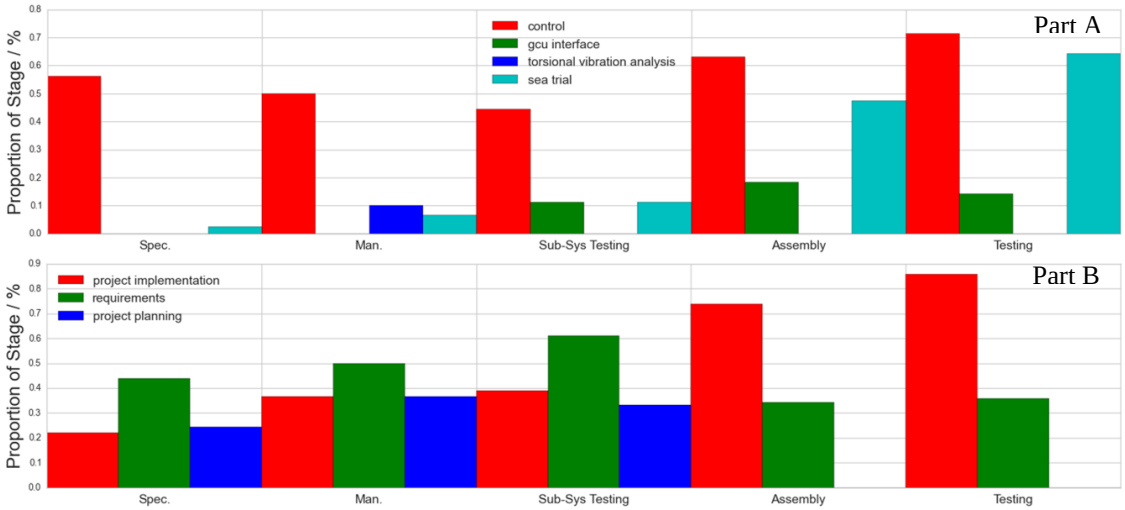

Fig 3: Persistence of 7 topics: Above - Control (red), GCU interface (green), torsional vibration analysis (blue), sea trial (cyan); below - project implementation (red), requirements (green), project planning (blue)

Figure 3 shows persistence by the process stages within which a term's episodes reside. This analysis can pick out features of individual topics and their appearance. For example, it demonstrates that the "GCU interface" (part A) was discussed occasionally during later-stages but not during early ones, the "sea trial" (part A) was important in later-stages but not early ones, as its discussion greatly increased through the process, and that "control" (part A) was a frequent topic throughout the process, likely with regard to control of a number of different elements. Such analysis allows managers to understand the relation between topics and process stages, and the level of activity dedicated to, or expected for, each. This can be used as a checking measure, a classification measure, or to highlight potential issues. For example, in part B it can be seen that "project planning" occurs throughout the early stages, but not during assembly or testing. This could either be due to all elements being planned, or could highlight that planning in later-stages is insufficient. "Requirements" are discussed throughout, as would likely be desired, but with a lower frequency in later-stages. This may signify a lack of adequate reference to requirements or specifications that a manager may wish to rectify. Finally "project implementation" is discussed with increasing frequency as the process nears the implementation phases; this may be confirmation to a manager that preperation for implementation is proceeding as desired.

Examining persistent episodes in this way allows a more detailed understanding of the appearance of topics through the engineering process to be built which, when analysed by an experienced manager in context of their own projects, has potential to support management and monitoring processes.

The topics can also be analysed to identify those that are most persistent in a given stage, while being less persistent in others. This gives a more general understanding of work that occurs in each stage of the engineering process, and has the benefit of highlighting those topics which are not ubiquitos - ie

removing the names of companies and projects. Table 4 gives the most localised terms for each stage - those with episodes covering a higher proportion of a single stage while with a lower proportion of another.

Table 4: Persistent topics in each stage

Stage	No. of localised terms	Topics
Specification	0	Project planning, drawing, proposal, short circuit, spare part list, specification
Manufacture	26	Scheduled FDS issue, lv short circuit, diagram rev, cargo piping, EPS meeting, spray pipe, project planning, flow rate, cargo handling, valve, compressor
Sub-system Testing	51	Risk assessment form, hd compressor top, routine test, battery back, network switch boxes, serial link test, cold ironing, pressure transmitter, electrical systems
Assembly	35	HV cable termination, cargo fat, gas purging, blackout recovery, gas combustion unit, GCU interface, differential pressure, production schedule, test procedure
Testing	37	Signal light columns, ship building division, 20v secondary transformer, onboard test procedure, fuel gas line, sea trial, purging, freight, project implementation

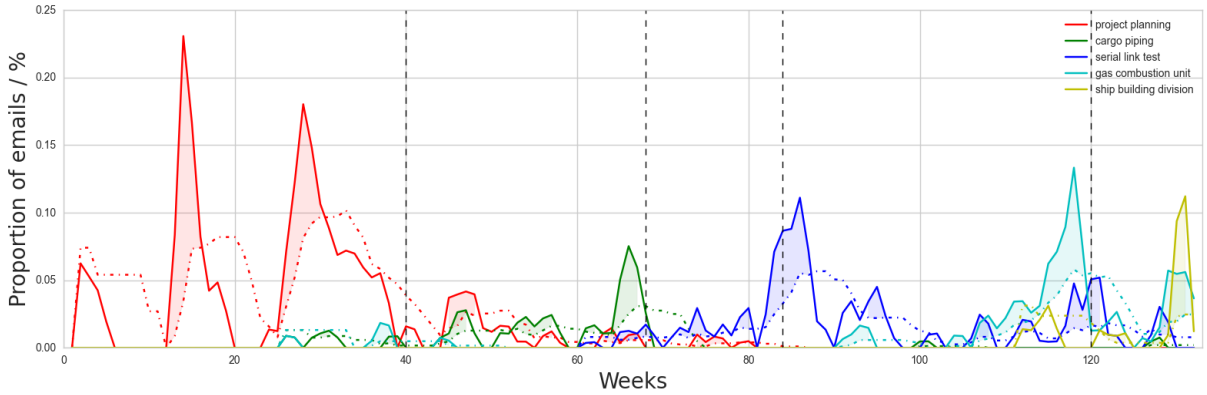

Fig 4: Patterns in locations of topic persistence

It is interesting in Table 4 that no topics were highly localised soley within the specification stage (ie. high appearance in specification in relation to other stages). This suggests that for this project all topics that appeared in the specification phase were carried through with continuing work to later stages. Depending on the project, this may be a positive or negative finding. The stage with most localised topics is sub-system testing, suggesting that the variety in topics of work was highest in this stage for this project. This may suggest to a manager that they should more closely monitor resources, or streamline staffing to quickly resolve work in certain areas. The localised terms also provide a summary of the specific work that occurred in each stage, that can be used as an historical record for monitoring of future projects, or as a check to observe if work is occurring according to the schedule.
This information, and that which could be interpreted by a manager working within the project context, aids in analysing the detail of the work that occurs in each process stage. Through highlighting the detail of the appearance of many topics of work and their place within the engineering process, topic persistence and localised persistence are able to provide information that may aid management processes.

5. Discussion

The detection of topic occurrence and persistence provides the opportunity to quickly generate information about engineering projects in real time, and hence aid managers in their work and decision making process. It provides insight into engineering projects automatically using a tangible and direct output of the project itself, and is real-time, extendable in scale, and not subject to limitations of such techniques as surveys and interviews, which can be hindered by difficulty in implementation, interpretation of results, and bias of respondents when employed for an application as presented here.

This therefore supports project monitoring in larger-scale projects - automatically determining what is occurring throughout and in specific process stages. The analysis presented here has multiple potential applications. Extending from the provision of information to a manager in real-time, the analysis of multiple projects over time gives the ability to compare and contrast historical cases, develop the general case of activity throughout a project in a given company, and support lessons learned.

It is not, however, and should not be, a purely automatic process. Throughout the analysis presented here, there has been the requirement for an experienced manager to interrogate and analyse the data within their context - first in selecting topics, second in selecting current and prior usage windows, and finally in interpreting results. Analysis enables quantified results but, due to the variation in project context and potential for indicators to have different meanings on a case-by-case basis, there is a need for managers to reason about project progress in a qualitative manner. As a result, while data analysis can be quantifiable and general, interpretation should be guided by theoretical considerations, with the sense-making and holistic understanding of projects generated by reflective interpretation of results (Thoring et al. 2015). The approach supports the manager in this subjective interpretation, and encourages their input through control over the analysis and selection of values in usage calculations.

As a manager must always interpret data, as opposed to automatic interpretation during analysis, there is a need for the approach to generate low-level and granular data - any summarisation and classification has potential to obscure information that may be important in a given context. However, there is signficant potential for summarisation to also aid understanding, particularly when used within research. For example, through classification of topics as managerial or technical, this approach has potential to provide an automatic method and extension to research on engineering activity that occurs throughout design research (see Wasiak et al. 2010). This is a valuable subject for further work.

The approach is also particularly suited to an organisational context. While the only data source employed here is email any textual data can be analysed - including reports, presentations, and other documents. This greater range of data can be analysed automatically using the same techniques, and would increase the level of detail and reliability in results. In addition, that the approach does not require any input from workers greatly reduces barriers to implementation that techniques such as self-reporting and diary studies often face, particularly for longer term studies.

In addition to the analyses presented here, there are many alternative opportunities for future work. Certain patterns, such as periodicity, sporadic persistence, high occurrence, and sudden occurrence, have all been identified as potentially interesting. Their detail has not, however, been explored. Through specific study of these features and their implication it may be possible to greatly increase the utility of analysis to managers through, for example, indication of the state of the project without their interpretation, or prediction of future events and issues. This is to be explored in future work.

Linked to the benefits of automatic analysis come the issue of validity when employing a naïve algorithm to perform analysis. While a human is needed to parse and interpret, those patterns highlighted here still require validation before they can be relied upon wholly. For example, while the topics found are important, it cannot be assumed that other important topics have not been omitted by the detection method, or that certain topics are given over or under-inflated importance by topic persistence measurement. Validation through interview or triangulation with other data, such as analyses using other digital assets, remains an important element of future work.

6. Conclusion

This paper has presented an approach to developing understanding of engineering projects through detection of email topic occurrence and persistence. The specific patterns identified in output data have potential to support the understanding and decision making processes of managers, giving greater detail of worker activity throughout the project process and in specific stages, highlighting potential issues or atypical occurrences, identifying core and peripheral areas of work, and monitoring the changing focus of work through the project lifespan. Through analysis of a long-term email corpus from a single project, this work has demonstrated the approach and its applicability, and the understanding to be gained from interpretation of an experienced manager. In addition, the approach has potential to provide a new data gathering method for research, that provides accurate, long-term, and reliable data of engineering activity.

References

Bendoly, E., 2014. System Dynamics Understanding in Projects: Information Sharing, Psychological Safety, and Performance Effects. Production and Operations Management, 23(December), pp.1352–1369.

Cash, P.J., Hicks, B.J. & Culley, S.J., 2013. A comparison of designer activity using core design situations in the laboratory and practice. Design Studies.

Chapman, C. & Ward, S., 1996. Project risk management: processes, techniques and insights.

Earl, C., Eckert, C. & Clarkson, J., 2005. Design Change and Complexity. In 2nd Workshop on Complexity in Design and Engineering. Glasgow, Scotland.

Durand, R., Decker, P.J. & Kirkman, D.M., 2014. Evaluation Methodologies for Estimating the Likelihood of Program Implementation Failure. American Journal of Evaluation, 35(3), pp.404-418.

Engwall, M., 2002. No project is an island : linking projects to history and context. Research Policy, 32(2003), pp.789–808.

Floricel, S. & Miller, R., 2001. Strategizing for anticipated risks and turbulence in large-scale engineering projects. International Journal of Project Management, 19(8), pp.445–455.

Gopsill, J.A. et al., 2015. Modelling the Evolution of Computer Aided Design Models : Investigating the Potential for Supporting Engineering Project Management. In PLM15: The 12th International Conference on Product Lifecycle Management. Doha, Qatar.

Gruhl, D. et al., 2004. Information diffusion through blogspace. ACM SIGKDD Explorations Newsletter, 6, pp.43–52.

Hales, C., 1987. Analysis of the Engineering Design Process in an Industrial Context. Cambridge: University of Cambridge.

Hill, A. et al., 2001. Identifying Shared Understanding in Design Using Document Analysis. In Proceedings of the 13th International Conference on Design Theory and Methodology. Pittsburgh, PA.

Jones, S. et al., 2015. Subject Lines As Sensors : Co-Word Analysis Of Email To Support The Management Of Collaborative Engineering Work. In ICED'15: International Conference on Engineering Design.

Luhn, H.P., 1957. A Statistical Approach to Mechanized Encoding and Searching of Literary Information. IBM Journal of Research and Development, 1(October), pp.309–317.

O'Donnell, F.J. & Duffy, A.H.B., 2002. Modelling design development performance. International Journal of Operations & Production Management, 22(11), pp.1198–1221.

O'Kane, P. & Hargie, O., 2007. Intentional and unintentional consequences of substituting face-to-face interaction with e-mail: An employee-based perspective. Interacting with Computers, 19(1), pp.20-31.

Robinson, M.A., 2010. An empirical analysis of engineers' information behaviours. Journal of the American Society for Information Science and Technology, 61(4), pp.640–658.

Snider, C. et al., 2015. Understanding Engineering Projects: An Integrated Vehicle Health Management Approach to Engineering Project Monitoring. In ICED15: International Conference on Engineering Design. Milan, Italy.

Stempfle, J. & Badke-Schaub, P., 2002. Thinking in design teams-an analysis of team communication. Design Studies, 23(5), pp.473–496.

Toor, S.-R. & Ogunlana, S.O., 2010. Beyond the "iron triangle": Stakeholder perception of key performance indicators (KPIs) for large-scale public sector development projects. International Journal of Project Management, 28(3), pp.228–236.

Thoring, K., Mueller, R.M. & Badke-Schaub, P., 2015. Technology-supported design research. In DS 80-11 Proceedings of the 20th International Conference on Engineering Design (ICED 15) Vol 11: Human Behaviour in Design, Design Education; Milan, Italy, 27-30.07. 15.

Wasiak, J. et al., 2010. Understanding engineering email: The development of a taxonomy for identifying and classifying engineering work. Research in engineering design, 21(1), pp.43–64.

Watson, J., 2012. Keynote address at the University of Bath.

Whittaker, S., Bellotti, V. & Gwizdka, J., 2006. Email in personal information management. Communications of the ACM, 49(1), p.68.

Wynn, D.C. and Clarkson, P.J., 2009. Design project planning, monitoring and re-planning through process simulation. In DS 58-1: Proceedings of ICED 09, the 17th International Conference on Engineering Design, Vol. 1, Design Processes, Palo Alto, CA, USA, 24.-27.08. 2009.

Xia, W. & Lee, G., 2004. Grasping the complexity of IS development projects. Communications of the ACM, 47, pp.68–74.